

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

Sub A
B1
1. Method for balancing the load on a plurality of servers that provide access to resources associated with a domain name, comprising:

(a) receiving a request for access to resources associated with the domain name;

(b) determining the load for each of a plurality of servers that provide access to resources associated with the domain name and selecting one of the plurality of servers to provide the access, the selection of the server being based on a determination for optimally balancing the load on the plurality of servers; and

(c) resolving an Internet protocol (ip) address of the selected server so that the accessing of resources associated with the domain name at the resolved ip address of the selected server will cause the load to be optimally balanced on the plurality of servers on a network.

2. The method of Claim 1, further comprises querying a local Domain Name System (DNS) to provide the ip address associated with the domain name.

3. The method of Claim 2, wherein when the ip address is not present at the local DNS, querying a primary DNS to resolve the ip address associated with the domain name.

4. The method of Claim 3, wherein when the primary DNS determines the domain name is delegated to a EDNS, further comprising referring the local DNS to the EDNS to resolve the ip address for the selected server, the EDNS employs at least one of a plurality of load balancing determinations to select one of the plurality of servers and resolve the ip address for the selected server.

5. The method of Claim 4, wherein the EDNS includes the primary DNS.

6. The method of Claim 4, wherein the EDNS includes a secondary DNS.

7. The method of Claim 4, wherein the EDNS is a primary EDNS, the primary EDNS collecting metric information employed by the selected load

balancing determination to select the server to provide access to the resources associated with the domain name.

8. The method of Claim 4, wherein selecting one of the plurality of servers that will optimally balance the load, further comprises choosing the server based on one of a plurality of static load balancing determinations for each server, the plurality of static load balancing determinations being selectable and including random, round robin, static ratio, global availability and topology.

9. The method of Claim 4, wherein selecting one of the plurality of servers that will optimally balance the load, further comprises choosing the server based on one of a plurality of dynamic load balancing determinations for each server, the dynamic load balancing determinations being selectable and including completion rate, least connections, packet rate, hops, round trip times, quality of service and dynamic ratio.

10. The method of Claim 4, further comprising selecting one of the plurality of load balancing determinations as a primary load balancing determination, the primary load balancing determination being used to select the server when a time stamp is not expired, the time stamp being associated with metric information used by the primary load balancing determination.

11. The method of Claim 10, further comprising selecting one of the plurality of load balancing determinations as an alternate load balancing determination, the alternate load balancing determination being employed to select the server when the time stamp associated with the metric information used by the primary load balancing determination is expired, another time stamp being associated with metric information employed by the alternate load balancing determination.

12. The method of Claim 11, further comprising selecting one of the plurality of load balancing determinations as a fallback load balancing determination, the fallback load balancing determination being employed to select the server when the time stamp associated with metric information used by the primary load balancing determination and the other time stamp associated with metric information employed by the alternate load balancing determination are expired.

Sub A³ 7

13. The method of Claim 7, further comprising a plurality of EDNSs that are separately disposed at a plurality of geographically distributed data centers, each data center including at least one of a server array controller, host machine and EDNS.

14. The method of 13, wherein at least one of the plurality of EDNSs is a secondary EDNS, the secondary EDNS storing a copy of the metric information collected by the primary EDNS, the secondary EDNS employing the copy of metric information to select a particular server at that will optimally balance the load for accessing resources.

15. The method of 13, wherein at least one of the plurality of EDNSs is a secondary EDNS, the secondary EDNS collecting metric information that is employed to select a particular server that will optimally balance the load for accessing resources

16. The method of Claim 4, further comprising a server array controller for managing access to at least one of the plurality of servers, the server array controller being in communication with the EDNS.

17. The method of Claim 16, wherein the server array controller is a BIG/IP server array controller.

18. The method of Claim 1, wherein the selected server is a stand-alone server.

19. The method of Claim 4, further comprising an agent program that collects the metric information and communicates the collected metric information to the EDNS when the EDNS is not resolving the ip address for the resources associated with the domain name request.

20. The method of Claim 1, wherein the network comprises a wide area network, Internet and intranet.

21. The method of Claim 4, further comprising a wide ip that maps the domain name to at least one server, the wide ip being employed when the primary DNS is separate from the EDNS.

22. The method of Claim 21, wherein the wide ip maps the domain name to one of the plurality of load balancing determinations.

23. The method of Claim 4, further comprising generating statistics from metric information collected by the EDNS and enabling statistics to be generated for a particular aspect of the network, including server array controller, host machine, server, path and wide ip configuration.

24. The method of Claim 23, wherein the load balancing determination is at least partially based on the generated statistics.

25. The method of Claim 23, wherein the statistics for the server array controller include the up versus down availability of the server array controller, number of packets between the EDNS and the server array controller, the total number of packets in and out of the server array controller, the number of packets processed by the kernel per second, the number of servers managed by the server array controller, the number of times data is refreshed, the amount of time the server array controller is active.

26. The method of Claim 23, wherein the statistics for the host machine include the number of servers managed by the host machine, number of times a particular host machine was chosen by a wide ip for load balancing and the number of times data is refreshed.

27. The method of Claim 23, wherein the statistics for the server include the number of times a particular machine was chosen by a wide ip for load balancing, the number of times data is refreshed, the number of connections that are handled by the server and the up versus down availability of the server.

28. The method of Claim 23, wherein the statistics for the path include the average round trip time (RTT) for transactions between the server array controller and a local DNS, the packet completion rate between the server array controller and the local DNS, the number of times a specified path is chosen, the number of times that the EDNS has received data about the specified path and the number of hops between routers for a transaction between the local DNS and the selected server.

29. The method of Claim 23, wherein the statistics for the local DNS include a measure of how often a particular local DNS is used and the number of times that the EDNS received a resolution request from the local DNS.

30. The method of Claim 22, wherein the statistics for the wide ip include weighting values for the servers managed by a particular server array controller, weighting values for the servers managed by another Host machine, the number of successful domain name resolutions, the number of unsuccessful name resolutions, the load balancing modes used for the pool of servers managed by each server array controller, the load balancing modes used for the pool of servers managed by each Host machine, the number of servers managed by each server array controller that are used to load balance a specified wide ip, and the number of servers managed by each host machine that are used to load balance the specified wide ip.

31. The method of Claim 4, wherein the EDNS employs an iQuery protocol to communicate the metric information from the agent program to the EDNS.

32. The method of Claim 1, wherein the EDNS is a 3DNS server.

33. The method of Claim 1, wherein at least a portion of the plurality of servers are virtual servers.

34. The method of Claim 24, wherein the generated statistics include a quality of service value that is related to the sum of separate portions of collected metric information, including packet rate, round trip time, hops, virtual server capacity, completion rate and topology.

35. The method of Claim 34, wherein each portion of the metric information is separately multiplied by a selectable value that determines the weight of that portion of the metric information in generating the quality of service value.

36. The method of Claim 34, wherein the generated statistics include a dynamic ratio value for each virtual server managed by a server array controller, the dynamic ratio value being related to the quality of service value and having selectable values for determining the weight of each portion of the metric information that is employed to generate the dynamic ratio value.

Sub A47

37. A system for balancing the load on a plurality of virtual servers that provide access to resources associated with a domain name, comprising:

(a) a memory for storing logical instructions; and
(b) a processor for executing the logical instructions stored in the memory, the execution of the logical instructions causing functions to be performed, including:

(i) receiving a request for access to resources associated with the domain name;

(ii) determining the load for each of a plurality of virtual servers that provide access to resources associated with the domain name and selecting one of the plurality of virtual servers to provide the access, the selection of the virtual server being based on a determination for optimally balancing the load on the plurality of virtual servers; and

(iii) resolving an Internet protocol (ip) address of the selected virtual server so that the accessing of resources associated with the domain name at the resolved ip address of the selected virtual server by the client will optimally balance the load on all of the virtual servers on a network.

38. Method for balancing the load on a plurality of virtual servers that provide access to resources associated with a domain name, comprising:

(a) receiving a request from a client for access to resources associated with the domain name;

(b) querying a local Domain Name System (DNS) to provide the Internet protocol (ip) address associated with the domain name;

(c) when the ip address is not present at the local DNS, querying a primary DNS to resolve the ip address associated with the domain name;

(d) when the primary DNS determines the domain name is delegated to a EDNS system, referring the local DNS to the EDNS system to resolve the ip address associated with the domain name; and

(e) employing the EDNS system to balance the load on a plurality of virtual servers that provide access to resources associated with the domain name by selecting one of the plurality of virtual servers that optimally balances the load, the EDNS system resolving the ip address of the selected virtual server for the domain name and providing the ip address to the client, so that the client will access

resources associated with the domain name at the resolved ip address of the selected virtual server.

39. A computer readable medium having computer executable instructions for performing the method recited in Claims 1, 4, 19 or 23.

Add A5 >